WILEY PUBLICATIONS IN STATISTICS

Contents

CHAPTER 1

Introduction

Statistical methods are essentially methods for dealing with data that have been obtained by a repetitive operation. For some sets of data, the operation that gave rise to the data is clearly of this repetitive type. This would be true, for example, of a set of diameters of a certain part in a mass-production manufacturing process or a set of percentages obtained from routine chemical analyses. For other sets of data, the actual operation may not seem to be repetitive, but it may be possible to conceive of it as being so. This would be true for the ages at death of certain insurance-policy holders, or for the total number of mistakes an experimental set of animals made the first time they ran a maze.

Experience indicates that many repetitive operations or experiments behave as though they occurred under essentially stable circumstances. Games of chance, such as coin tossing or dice rolling, usually exhibit this property. Many experiments and operations in the various branches of science and industry do likewise. Under such circumstances, it is often possible to construct a satisfactory mathematical model of the repetitive operation. This model can then be employed to study properties of the operation and to draw conclusions concerning it. Although mathematical models are especially useful devices for studying real-life problems when the model is realistic of the actual operation involved, it often happens that such models prove useful even though the operation is not highly stable.

The mathematical model that a statistician selects for a repetitive operation is usually one that enables him to make predictions about the frequency with which certain results can be expected to occur when the operation is repeated a number of times. For example, the model for studying the inheritance of color in the propagation of certain flowers might be one that predicted 3 times as many flowers of one color as of another color. In the investigation of the quality of manufactured parts, the model might be one that predicts the percentage of defective parts that can be expected in the manufacturing process.

Because of the nature of statistical data and models, it is only natural that probability should be the fundamental tool in statistical theory. The statistician looks upon probability as an idealization of the proportion of times that a certain result will occur in repeated trials of an experiment; consequently a probability model is the type of mathematical model selected by him. Because probability is so important in the theory and applications of statistical methods, a brief introduction to probability will be given before the study of statistical methods as such is taken up.

The idea of a mathematical model for assisting in the solution of real-life problems is a familiar one in the various sciences. For example, a physicist studying projectile motion often assumes that the simple laws of mechanics yield a satisfactory model, in spite of the complexity of the actual problem. For more refined work, he introduces a more complicated model. Since a model is only an idealization of the actual situation, the conclusions derived from it can be relied upon only to the extent that the model chosen is a sufficiently good approximation to the actual situation being studied. In any given problem, therefore, it is essential to be well acquainted with the field of application in order to know what models are likely to be realistic. This is just as true for statistical models as for models in the various branches of science.

The science student will soon discover the similarity between certain of the statistical methods and certain scientific methods in which the scientist sets up a hypothesis, conducts an experiment, and then tests the hypothesis by means of his experimental data. Although statistical methods are applicable to all branches of science they have been applied most actively in the biological and social sciences, because the laboratory methods of the physical sciences have not been sufficiently broad to treat many of the problems of those other sciences. Problems in the biological and social sciences often involve undesired variables that cannot be controlled, as contrasted to the physical sciences in which such variables can often be controlled satisfactorily in the laboratory. Statistical theory is concerned not only with how to solve certain problems of the various sciences, but it is also concerned with how experiments in those sciences should be designed. Thus, the science student should expect to learn statistical techniques not only to assist him in treating his experimental data but also to assist him in designing his experiments in a more efficient manner.

The theory of statistics can be treated as a branch of mathematics in which probability is the basic tool; however, since the theory developed from an attempt to solve real-life problems, much of the theory would

not be fully appreciated if it were removed from such applications. Therefore the theory and the applications will be considered simultaneously throughout this book, although the emphasis will be on the theory.

In the process of solving a real-life problem in statistics, three steps may be recognized. First, a mathematical model is selected. Second, a check is made as to the reasonableness of the model. Third, the proper conclusions are drawn from this model to solve the proposed problem. In this book, the emphasis will be on the first and third steps. In order to do justice to the second step it would be necessary to be well acquainted with the field of application. It would also be necessary to know how the conclusions are affected by changes in the assumptions necessary for the model.

Students who have not had experience with applied science are sometimes disturbed by the readiness with which a statistician will accept certain of his model assumptions as being sufficiently well satisfied in a given problem to justify confidence in the validity of the conclusions. One of the striking features of much of statistical theory is that its field of application is much broader than the assumptions involved would seem to justify. The rapid development of, and interest in, statistical methods during the past few decades can be attributed in part to the highly successful application of statistical techniques to so many different branches of science and industry.

REFERENCES

A fuller discussion of some of the preceding ideas may be found in the following books:

Neyman, J., *First Course in Probability and Statistics*, Henry Holt and Co., pp. 1–6.

Fisher, R. A., *Statistical Methods for Research Workers*, Oliver & Boyd, Chapter 1.

Kendall, M. G., *The Advanced Theory of Statistics*, Griffin and Co., pp. 164–166.

Wilks, S. S., *Mathematical Statistics*, Princeton University Press, pp. 1–4.

CHAPTER 2

Probability

2.1 Introduction

An individual's approach to probability depends upon the nature of his interest in the subject. The pure mathematician usually prefers to treat probability from an axiomatic point of view just as he does, say, the study of geometry. The applied statistician usually prefers to think of probability as the proportion of times that a certain event will occur if the experiment related to the event is repeated indefinitely. The approach to probability here will be based on a blending of these two points of view.

The statistician is interested in probability only as it pertains to the possible outcomes of experiments. Furthermore, he is interested in only those experiments that are repetitive in nature, or that can be conceived of as being so. Experiments such as tossing a coin, counting the number of defective parts in a box of parts, or reading the daily temperature on a thermometer are examples of simple repetitive experiments. An experiment in which several experimental animals are fed different rations in an attempt to determine the relative growth properties of the rations may be performed only once with those same animals; nevertheless the experiment may be thought of as the first in an unlimited number of similar experiments and therefore it may be conceived of as being repetitive.

2.2 Sample Space

Consider a simple experiment such as tossing a coin. In this experiment there are but 2 possible outcomes, a head or a tail. It is convenient to represent the possible outcomes of such an experiment, and experiments in general, by points on a line, or by points in higher dimensions. Here it would be convenient to represent a head by the point 1 on the x axis and a tail by the point 0. This choice is convenient because the number corresponds to the number of heads obtained in the toss. If the experiment had consisted of tossing the coin twice,

there would have been 4 possible outcomes, namely HH, HT, TH, TT. For reasons of symmetry, it would be desirable to represent these 4 possible outcomes by the points (1, 1), (1, 0), (0, 1), and (0, 0) in the xy plane. Figure 1 illustrates this choice of points to represent the possible outcomes of the experiment.

If the coin were tossed 3 times, it would be convenient to use 3 dimensions to represent the possible experimental outcomes. This representation, of course, is merely a convenience, and if desired one could just as well mark off any 8 points on the x axis to represent the 8 possible outcomes.

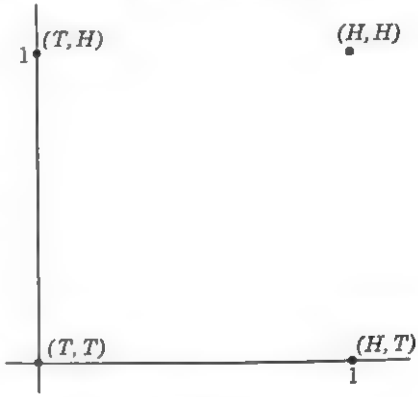

FIG. 1. A simple sample space.

DEFINITION: *The set of points representing the possible outcomes of an experiment is called the sample space, or the event space, of the experiment.*

The idea of a sample space is introduced because it is a convenient mathematical device for developing the theory of probability as it pertains to the outcomes of experiments.

2.3 Sample Space Probabilities

Experience with experiments shows that for some experiments one possible outcome is much more likely to occur than another possible outcome. For example, in counting the number of defective screws in a box of screws purchased from a reputable firm, one is much more likely to find all good screws than all defective screws. In many simple games of chance, however, it often happens that all the possible outcomes will occur about equally often in a large number of repetitions of the experiment. Thus, in tossing a die repeatedly, each of the 6 sides will usually occur with about the same frequency.

Before it is possible to discuss the probability of some combination of possible experimental outcomes, it is necessary that probabilities be assigned to each of the sample points in the sample space. Since the interpretation of probability is going to be in terms of frequency, the probability that is assigned to a given sample point should be approximately equal to the proportion of times that the sample point will be obtained, or is expected to be obtained, in a large number of repetitions of the experiment. This frequency interpretation of probability requires that probabilities be non-negative and that the sum of the

probabilities assigned to the sample points be equal to one; hence, probabilities must be assigned with this restriction in mind. In the preceding illustration of tossing a coin twice, it would be natural to assign the probability of $\frac{1}{4}$ to each of the 4 sample points, unless experience has indicated that the coin is biased, that is, that one side comes up more frequently than the other. The assignment of probabilities to each of the possible outcomes in sampling a box of screws for defectives would need to be based on experience with the manufacturer's product. From a mathematical point of view, any set of non-negative numbers totaling one may be assigned to the sample points as probabilities; however, the conclusions derived from the theory are not likely to prove very realistic unless the sample-point probabilities are chosen in a realistic manner. The assignment of probabilities to the sample points constitutes the first step in the process of choosing a mathematical model for the real-life experiment under consideration.

Since the development of the theory of probability is especially simple when there are only a finite number of sample points and when all sample points are assigned the same probability, it will be assumed in the next few sections that the sample space is of this simple type. Many games of chance possess sample spaces that are naturally of this type. Thus, in rolling a die twice, it is natural to assign equal probabilities ($\frac{1}{36}$) to the 36 sample points that constitute the sample space.

2.4 Events

Consider an experiment such that whatever the outcome of the experiment, it can be decided whether or not an event A has occurred. This means that each sample point can be classified as one for which A will occur, or as one for which A will not occur. Now the general definition of the probability that an event A will occur is usually given as being the sum of the probabilities of the sample points corresponding to the occurrence of A. However, since the discussion is being limited to finite sample spaces with equal sample point probabilities, the general definition can be simplified as follows.

(1) DEFINITION: *The probability that an event A will occur is the ratio of the number of sample points that correspond to the occurrence of A to the total number of sample points.*

In symbols, if $P\{A\}$ denotes the probability that A will occur when the experiment is performed and if $n(A)$ and n denote the number of sample points giving rise to A and the total number of sample points,

respectively, then

$$P\{A\} = \frac{n(A)}{n} \tag{2}$$

As an illustration, suppose a coin is tossed twice and suppose that all 4 sample points are assigned the same probability. Then the probability of getting a total of 1 head and 1 tail is $2/4$ because the 2 sample points (1, 0) and (0, 1) correspond to the occurrence of the desired event. In rolling a die twice, if it is assumed that all 36 sample points are assigned the same probability, the probability that the sum of the face numbers will be 7 is $6/36$ because the sample points (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), and (6, 1) correspond to the occurrence of the desired event.

In the illustrations as well as in the theory of the next few sections, the assumption that the sample points are assigned the same probability will not be stated explicitly each time as was done in the preceding illustrations.

2.5 Addition Theorem

Applications of probability are often concerned with a number of related events rather than with just 1 event. For simplicity, consider 2 such events, A_1 and A_2, associated with an experiment. One may be interested in knowing whether both A_1 and A_2 will occur when the experiment is performed. This joint event will be denoted by the product A_1A_2 and its probability by $P\{A_1A_2\}$. On the other hand, one may be interested in knowing whether at least 1 of the events A_1 and A_2 will occur when the experiment is performed. This event will be denoted by the sum $A_1 + A_2$ and its probability by $P\{A_1 + A_2\}$. At least 1 of the 2 events will occur if A_1 occurs but A_2 does not, or if A_2 occurs but A_1 does not, or if both A_1 and A_2 occur. The purpose of this section is to derive a formula for $P\{A_1 + A_2\}$.

Let the sample space for an experiment be represented by the points in Fig. 2 and let the sample points corresponding to the occurrence of A_1 and A_2 be the points interior to the regions labeled A_1 and A_2, respectively. The points common to these 2 regions determine a region that has been labeled A_1A_2. This notation makes it clear that the region A_1A_2 is part of the region A_1 and also part of the region A_2.

From definition (1), it follows that $P\{A_1 + A_2\}$ is the ratio of the number of sample points lying inside the two regions A_1 and A_2 combined, to the total number of sample points. But the number of sample points lying inside the two regions A_1 and A_2 is equal to the number lying inside region A_1, plus the number lying inside region A_2,

minus the number lying inside the common region A_1A_2, because the points lying inside A_1A_2 would be counted twice if no subtraction were made. This counting can be written symbolically as

$$n(A_1 + A_2) = n(A_1) + n(A_2) - n(A_1A_2)$$

If both sides of this equation are divided by n, the total number of

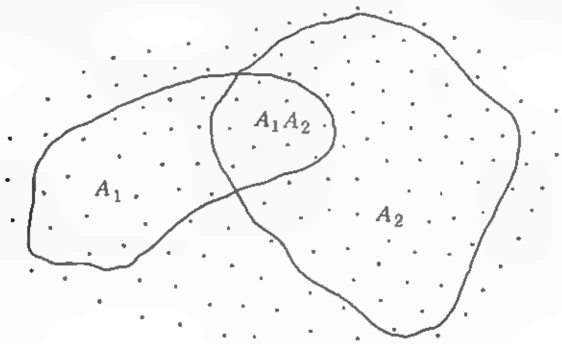

FIG. 2. A general sample space.

sample points, and if definition (1) is applied, the following fundamental theorem, known as the addition theorem, will be obtained.

(3) ADDITION THEOREM:

$$P\{A_1 + A_2\} = P\{A_1\} + P\{A_2\} - P\{A_1A_2\}$$

Two events A_1 and A_2 often have no sample points in common. When this occurs, the events A_1 and A_2 are said to be mutually exclusive because if one of the events occurs the other cannot occur. Formula (3) then reduces to the following formula:

(4) $$P\{A_1 + A_2\} = P\{A_1\} + P\{A_2\} \text{ \textit{when} } A_1 \text{ \textit{and} } A_2 \text{ \textit{are mutually exclusive}}$$

Formulas (3) and (4) can be generalized to more than 2 events. The generalization of (4) is obvious and will be used in later work. The generalization of (3) is more complicated; however since the generalization will not be needed in later work, it will not be considered here.

2.6 Multiplication Theorem

The purpose of this section is to derive a formula for $P\{A_1A_2\}$ in terms of probabilities of single events. In order to do so, it is necessary to introduce the notion of conditional probability. Suppose that

one is interested in knowing whether A_2 will occur, subject to the condition that A_1 is certain to occur. Since A_1 is certain to occur only when the sample space is restricted to those points lying inside the region labeled A_1 in Fig. 2, the total number of sample points is now reduced to $n(A_1)$. It will be assumed that $n(A_1) > 0$, that is, that at least 1 possible outcome of the experiment will correspond to the occurrence of A_1. Among these $n(A_1)$ points, the number corresponding to the occurrence of A_2 is $n(A_1A_2)$. If the probability that A_2 will occur subject to the restriction that A_1 is certain to occur is denoted by $P\{A_2 \mid A_1\}$, then it follows from definition (1) that

(5) $$P\{A_2 \mid A_1\} = \frac{n(A_1A_2)}{n(A_1)}$$

This probability is called the conditional probability of A_2, subject to the condition A_1. It is often called the probability that A_2 will occur, it being known that A_1 has occurred. Since

$$P\{A_1A_2\} = \frac{n(A_1A_2)}{n}$$

and

$$P\{A_1\} = \frac{n(A_1)}{n}$$

it follows from dividing the numerator and denominator on the right side of (5) by n that

(6) $$P\{A_2 \mid A_1\} = \frac{P\{A_1A_2\}}{P\{A_1\}}$$

This formula, when written in product form, yields the fundamental multiplication theorem for probabilities.

(7) Multiplication Theorem: $P\{A_1A_2\} = P\{A_1\}P\{A_2 \mid A_1\}$.

Formula (6) holds only when $P\{A_1\} \neq 0$. Formula (7), however, may be treated as holding in general if it is agreed to give the right side the value 0 when the factor $P\{A_1\}$ is equal to 0. If the order of the two events is interchanged, formula (7) becomes

(8) $$P\{A_1A_2\} = P\{A_2\}P\{A_1 \mid A_2\}$$

Now, suppose that A_1 and A_2 are 2 events such that $P\{A_2 \mid A_1\} = P\{A_2\}$ and such that $P\{A_1\}P\{A_2\} > 0$. Then the event A_2 is said to be independent in a probability sense, or more briefly, independent, of the event A_1. This name follows from the property that the proba-

bility of A_2 occurring is not affected by adding the condition that A_1 must occur. When A_2 is independent of A_1, (7) reduces to

$$P\{A_1A_2\} = P\{A_1\}P\{A_2\} \tag{9}$$

Conversely, when (9) is true, it follows from comparing (9) and (7) that A_2 is independent of A_1. If the right members of (8) and (9) are equated, it will be seen that $P\{A_1 \mid A_2\} = P\{A_1\}$. But this states that the event A_1 is independent of the event A_2. Thus, if A_2 is independent of A_1, it follows that A_1 must be independent of A_2. Because of this mutual independence and because (9) implies this independence, it is customary to define independence in the following manner.

(10) DEFINITION: *Two events, A_1 and A_2, are said to be independent if $P\{A_1A_2\} = P\{A_1\}P\{A_2\}$.*

Formulas (7) and (10) can be generalized in an obvious manner for more than two events by always combining events into two groups.

2.7 Illustrations

As illustrations of how the preceding rules of probability apply, consider a few card problems.

Two cards are drawn from an ordinary deck of cards, the first card drawn being replaced before the second card is drawn. (*a*) What is the probability that both cards will be spades? Here, A_1 denotes the event of getting a spade on the first draw and A_2 the event of getting a spade on the second draw. Since the card drawn is replaced each time, events A_1 and A_2 may be assumed to be independent; hence, formula (9) applies with the result that $P\{A_1A_2\} = {}^{13}\!/_{52} \cdot {}^{13}\!/_{52} = {}^{1}\!/_{16}$. (*b*) What is the probability that at least 1 of the 2 cards drawn will be a spade? Formula (3) applies here; hence, using the answer to part (*a*), $P\{A_1 + A_2\} = {}^{13}\!/_{52} + {}^{13}\!/_{52} - {}^{1}\!/_{16} = {}^{7}\!/_{16}$. This problem could also be solved indirectly by first calculating the probability that neither card drawn will be a spade, which is ${}^{39}\!/_{52} \cdot {}^{39}\!/_{52}$, and then subtracting this result from 1. The reasoning here is that the converse of "neither card will be a spade" is "at least 1 card will be a spade."

Two cards are drawn from a deck but the first card drawn is not replaced. (*c*) What is the probability that both cards will be spades? Formula (7) now applies; hence $P\{A_1A_2\} = {}^{13}\!/_{52} \cdot {}^{12}\!/_{51} = {}^{1}\!/_{17}$. (*d*) What is the probability that the second card drawn will be a spade? Here A_2 can occur only if 1 of the 2 mutually exclusive events A_1A_2 or $\bar{A}_1A_2$ occurs, where $\bar{A}_1$ denotes the nonoccurrence of A_1; hence, by formula (4), $P\{A_2\} = P\{A_1A_2\} + P\{\bar{A}_1A_2\}$. But from

(c), $P\{A_1A_2\} = 1/17$. Using formula (7), $P\{\bar{A}_1A_2\} = 39/52 \cdot 13/51 = 13/68$. Combining these results, $P\{A_2\} = 1/17 + 13/68 = 1/4$. This result shows what is intuitively clear, that the probability of getting a spade on the second draw when the result of the first draw is unknown is the same as the probability of getting a spade on the first draw.

2.8 Combinatorial Formulas

The simplest problems on which to develop facility in applying the addition and multiplication rules of probability are some of the problems related to games of chance. For many such problems, however, the counting of sample points corresponding to various events becomes tedious unless compact counting methods are developed. A few of the formulas that yield such methods will be derived in this section.

2.8.1 Permutations

Consider a set of n different objects, such as n blocks having different numbers or colors. Let r of the n objects be selected and arranged in a line. Such an arrangement is called a *permutation* of the r objects. If two of the r objects are interchanged in their respective positions, a different permutation results. In order to count the total number of such permutations, it suffices to consider the r positions on the line as fixed and then count the number of ways in which blocks can be selected to be placed in the r positions. Starting from the position farthest to the left, any one of the n blocks may be chosen to fill this position. After the first position has been filled, there will be but $n - 1$ blocks left to choose from to fill the second position. For each choice for the first position, there are therefore $n - 1$ choices for the second position; and hence $n(n - 1)$ total choices for the two positions. If this selection procedure is continued, there will be $n - r + 1$ blocks left to choose from for the rth position. If the total number of such permutations is denoted by ${}_nP_r$, it therefore follows that

$$(11) \qquad {}_nP_r = n(n-1)\cdots(n-r+1)$$

The symbol ${}_nP_r$ is usually called the number of permutations of n things taken r at a time.

As an illustration, suppose one is given the 4 letters a, b, c, d. The number of permutations of these 4 letters taken 2 at a time is given by ${}_4P_2 = 4 \cdot 3 = 12$. These permutations are easily enumerated as follows: *ab, ba, ac, ca, ad, da, bc, cb, bd, db, cd, dc*.

If r is chosen equal to n, (11) reduces to

$$(12) \qquad {}_nP_n = n(n-1)\cdots(1) = n!$$

In order to permit formulas that involve factorials to be correct even when $n = 0$, it is necessary to define $0! = 1$. This is consistent for $n = 1$ with the factorial property that $(n - 1)! = n!/n$.

2.8.2 Combinations

If one is interested only in what particular objects are selected, when r objects are chosen from n objects, without regard to their arrangement in a line, then the unordered selection is called a *combination*. Thus, if 2 letters are chosen from the 4 letters a, b, c, d, the combination ab is the same combination as ba, but of course differs from the combination ac. The total number of combinations possible in selecting r objects from n different objects will be denoted by the symbol $\binom{n}{r}$. This symbol is usually called the number of combinations of n things taken r at a time.

In order to derive a formula for $\binom{n}{r}$, it suffices to compare the total number of permutations and total number of combinations possible. Since a permutation is obtained by first selecting r objects and then arranging them in some order, whereas a combination is obtained by performing only the first step, it follows that the total number of permutations is obtained by taking every possible combination, the total number of which is $\binom{n}{r}$, and arranging them in all possible ways. But from (12) the total number of arrangements of r objects in r places is $r!$; hence, the total number of permutations is given by multiplying the number of combinations, $\binom{n}{r}$, by $r!$. Thus, ${}_nP_r = \binom{n}{r}\cdot r!$. Using formula (11), it therefore follows that

$$\binom{n}{r} = \frac{n(n-1)\cdots(n-r+1)}{r!} \tag{13}$$

Since $n(n-1)\cdots(n-r+1) = n!/(n-r)!$, formula (13) may be written in the following more compact form:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \tag{14}$$

As an illustration, the number of combinations of 2 letters selected from the 4 letters a, b, c, d is given by $\binom{4}{2} = 4!/2!2! = 6$. The actual combinations are ab, ac, ad, bc, bd, cd.

2.8.3 Permutations When Some Elements Are Alike

In the preceding derivations it has been assumed that all the n objects were different. It sometimes happens, however, that the n objects contain a number of similar objects. Thus, one might have 5 colored balls of which 3 were white and 2 were black, instead of 5 distinct colors. Now suppose that there are but k distinct kinds of objects and that there are n_1 of the first kind, n_2 of the second kind, $\cdots$, and n_k of the kth kind, where $n_1 + n_2 + \cdots + n_k = n$. The total number of different permutations of these n objects arranged in a line is obviously less than $n!$. In order to find the total number of distinct permutations, it suffices to compare the number of permutations now, which will be denoted by P, with the number that would be obtained if the like objects were given marks to distinguish them. The comparison is similar to that made between $\binom{n}{r}$ and ${}_nP_r$ in deriving formula (13). Each permutation in the problem under consideration gives rise to additional permutations when the like objects are made different by markings. For example, if the n_1 similar objects in a permutation are made different, they can be rearranged in their positions in $n_1!$ ways. Since this will be true for each of the P permutations, there will be $n_1!$ times as many permutations when the n_1 similar objects are made different as before. In the same manner, the n_2 similar objects may be made different to give $n_2!$ times as many permutations as before. Continuing this procedure, the total number of permutations after all similar objects have been made different will be $n_1!n_2!\cdots n_k!$ times as large as the number of permutations before the similar objects were made different; hence, the total number after these changes will be $Pn_1!n_2!\cdots n_k!$. But after all similar objects have been made different, the total number of permutations will be the number of permutations of n different things taken n at a time, which is $n!$. Equating these 2 results and solving for P, one obtains

$$\frac{n!}{n_1!n_2!\cdots n_k!} \tag{15}$$

for the total number of permutations of n things in which there are n_1 alike, n_2 alike, $\cdots$, n_k alike. As an illustration, consider the number of permutations of the 5 letters a, a, a, b, b. Formula (15) yields $5!/3!2! = 10$. These permutations are easily written down: *aaabb, aabab, abaab, baaab, aabba, ababa, baaba, abbaa, babaa, bbaaa.*

2.8.4 Illustrations of the Use of Combinatorial Formulas

(*a*) Consider a bridge hand consisting of 13 cards chosen from an ordinary deck. What is the probability that such a hand will contain exactly 7 spades? Since a bridge hand is not concerned with the order in which the various cards are obtained, the total number of possible bridge hands is equal to the number of ways of choosing 13 objects from 52 objects, or $\binom{52}{13}$. This is therefore the total number of sample points in the sample space. The number of hands containing exactly 7 spades is equal to the number of ways of choosing 7 spades from 13 spades, or $\binom{13}{7}$, multiplied by the number of ways of choosing 6 nonspades from 39 nonspades, or $\binom{39}{6}$. Hence, the desired probability is given by

$$\frac{\binom{13}{7}\binom{39}{6}}{\binom{52}{13}} = \frac{13!39!13!39!}{7!6!6!33!52!}$$

(*b*) If a coin is tossed 5 times, what is the probability that 3 heads and 2 tails will be obtained? First, consider a fixed order in which the desired result can occur, say $HHHTT$. From (10) the probability of obtaining this particular order of events is $(\frac{1}{2})^5$. Any other ordering of these 3 H's and 2 T's will have the same probability of being obtained. Next, consider the number of possible orderings. This number is equal to the number of permutations of 5 letters of which 3 are alike and 2 are alike, which by formula (15) is equal to $5!/3!2! = 10$. Since the 10 orderings constitute the mutually exclusive ways in which the desired event can occur, formula (4) yields the desired answer, namely, $10(\frac{1}{2})^5 = \frac{5}{16}$.

(*c*) A pair of coins is tossed 200 times. What is the probability that exactly x of the 200 tosses will show double heads? As in the preceding illustration, consider a fixed order in which the desired result can occur, say,

$$\overbrace{SS\cdots S}^{x}\ \overbrace{FF\cdots F}^{200-x}$$

where S denotes a success, that is, a double head, and F a failure, and where there are x successes and $200 - x$ failures. Because of the independence of the trials, the probability that this particular ordering

will be obtained is $(\frac{1}{4})^x(\frac{3}{4})^{200-x}$. The number of such orderings is equal to the number of permutations of the S's and F's, which in turn is equal to the number of permutations of 200 things of which x are alike and $200 - x$ are alike. By formula (15), this number is $200!/x!(200 - x)!$. Since these orderings constitute the mutually exclusive ways in which the desired event can occur, it follows that the desired probability is given by

$$(16) \qquad \frac{200!}{x!(200-x)!}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{200-x}$$

2.9 Random Variables

Consider a sample space corresponding to the tossing of two coins and suppose that interest is centered on the total number of heads that will be obtained. In order to study probabilities of such events, it is convenient to introduce a variable x that will give the total number of heads obtained. If the sample space suggested in (2.2) and displayed in Fig. 1 is used, the variable x will assume the value 0 at the sample point (0, 0), the value 1 at the sample points (1, 0) and (0, 1), and the value 2 at the sample point (1, 1). A numerical-valued variable x such as this is an example of what is called a random, or chance, variable.

(17) Definition: *A random variable is a numerical-valued variable defined on a sample space.*

As an illustration, if x denotes the sum of the points obtained in rolling 2 dice, then x is a random variable that can assume integral values from 2 to 12. The sample space here consists of 36 sample points. As another illustration, if 4 cards are drawn from a deck and if x denotes the number of black cards obtained, then x is a random variable that can assume integral values from 0 to 4. The sample space here consists of $\binom{52}{4}$ sample points.

The name random, or chance, is given to variables such as those in these illustrations because the variables are defined on sample spaces associated with physical experiments in which the outcome of any one experiment is uncertain and is therefore said to depend upon chance. Although the physical experiment that suggested the sample space is of this type, after the sample space has been chosen and a random variable x has been defined on it, the random variable x is just an ordinary variable of mathematics which can be assigned values over its range of values at pleasure.

2.10 Frequency Functions

After a random variable x has been defined on a sample space, interest usually centers on determining the probability that x will assume specified values in its range. From (1), the probability that x will assume a particular value, say x_0, is equal to the number of sample points for which $x = x_0$, divided by the total number of sample points. The relationship between the value of x and its probability is expressed by means of a function called the frequency function, which is defined as follows.

(18) DEFINITION: *A function $f(x)$ that yields the probability that the random variable x will assume any particular value in its range is called the frequency function of the random variable x.*

A frequency function often consists of merely a table of values. Thus, if 2 coins are tossed and if x denotes the total number of heads obtained, it suffices to define $f(x)$ by means of the following set of values: $f(0) = \frac{1}{4}, f(1) = \frac{1}{2}, f(2) = \frac{1}{4}$.

In the following chapters, when explicit mathematical models are selected for experiments several important frequency functions will be found that will be given by means of formulas rather than by tables of values. The function defined by (16) is an example of a frequency function defined by a formula.

In order to judge quickly how a variable is distributed, that is, how its probability changes as the variable changes, it is convenient to graph the frequency function $f(x)$ by means of a line graph. As an illustration of such a graph, let x denote the sum of the points obtained in rolling a pair of dice. Enumeration of cases will show that $f(2) = f(12) = \frac{1}{36}$, $f(3) = f(11) = \frac{2}{36}$, $f(4) = f(10) = \frac{3}{36}$, $f(5) = f(9) = \frac{4}{36}$, $f(6) = f(8) = \frac{5}{36}$, and $f(7) = \frac{6}{36}$. The line graph of $f(x)$ is given in Fig. 3.

A function closely related to the frequency function $f(x)$ is the *distribution function* $F(x)$. It is defined by the relation

(19)
$$F(x) = \sum_{t \leq x} f(t)$$

where the summation occurs over all those values of the random variable that are less than or equal to the specified value of x. Thus $F(x_0)$ gives the probability that the random variable x will assume a value less than or equal to x_0, as contrasted to $f(x_0)$ which gives the probability that x will assume the particular value x_0. The function $F(x)$ is called the distribution function by pure mathematicians but is

often called the cumulative distribution function by statisticians. This difference in terminology is due to the fact that pure mathematicians and statisticians have worked separately on closely related problems and have given different names to the same functions.

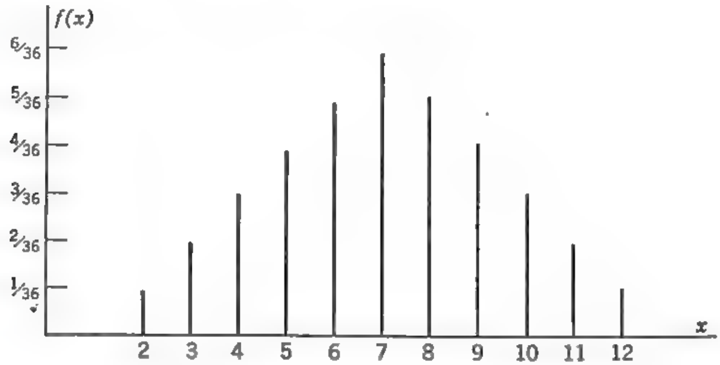

FIG. 3. Line graph for a frequency function.

Since some statisticians call $f(x)$ the distribution function, they naturally would call $F(x)$ the cumulative distribution function. Although this is a book for statisticians and written by a statistician, the terminology of the pure mathematician is being used in an attempt

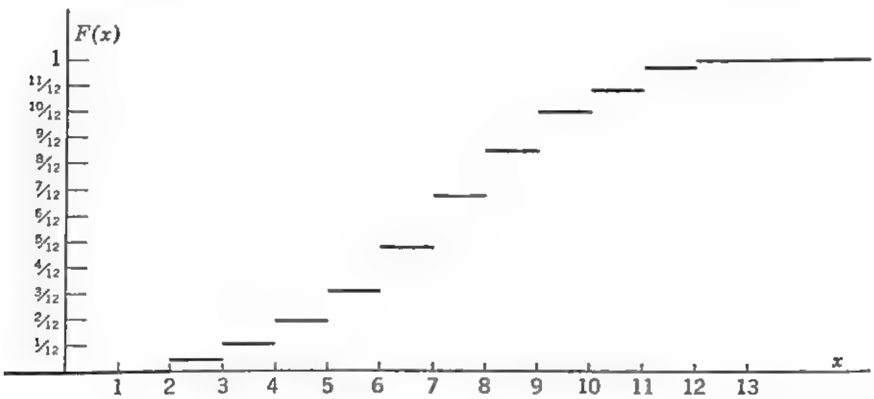

FIG. 4. Graph of a distribution function.

to foster a common terminology. The pure mathematician seldom reads statistical literature, whereas the statistician reads both statistical and mathematical literature. As a consequence, the pure mathematician is not likely to be affected by differences in terminology, whereas the statistician is continuously bothered by this dual termi-

nology, and it is therefore up to him to do the changing if he wishes to avoid being bothered.

The graph of $F(x)$ for the illustration of the preceding paragraph is given in Fig. 4. It should be noted that the value of $F(x)$ for x an integer is the upper value rather than the lower.

2.11 Joint Frequency Functions

Many experiments involve several random variables rather than just 1 such variable. For simplicity, consider 2 random variables x and y. A mathematical model for these 2 variables will be a function that will give the probability that x will assume a particular value while at

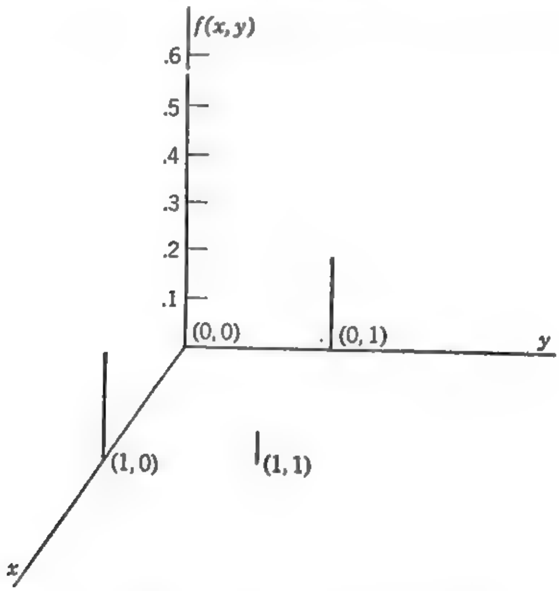

FIG. 5. Graph of a joint frequency function.

the same time y will assume a particular value. A function $f(x, y)$ that gives such probabilities is called a *joint frequency function* of the 2 random variables x and y.

As an illustration, let x denote the number of spades obtained in drawing one card from an ordinary deck and let y denote the number of spades obtained in drawing a second card from the deck, without the first card being replaced. Then $f(x, y)$ will be defined by the following table of values: $f(0, 0) = {}^{39}\!/_{52} \cdot {}^{38}\!/_{51}$; $f(1, 0) = {}^{13}\!/_{52} \cdot {}^{39}\!/_{51}$; $f(0, 1) = {}^{39}\!/_{52} \cdot {}^{13}\!/_{51}$; and $f(1, 1) = {}^{13}\!/_{52} \cdot {}^{12}\!/_{51}$. The graph of $f(x, y)$ as a line graph is given in Fig. 5.

In much of the statistical theory that will be developed in the

following chapters, the variables will be unrelated in a probability sense. In the preceding illustration, the variables x and y would have been such variables if the first card drawn had been replaced before the second card was drawn. To say that variables are unrelated in a probability sense means that the probability of one of the variables assuming a particular value is independent of what values the other variables assume. Random variables possessing this property are said to be independently distributed and are called independent random variables. In order to define independence more precisely, let $f(x_1, x_2, \cdots, x_n)$ be the joint frequency function of the indicated variables and let $f_i(x_i)$ denote the frequency function of the variable x_i. Then the essential property of such variables follows from the definition of independent events given by (10) and may be formalized in the following manner.

(20) DEFINITION: *If the joint frequency function $f(x_1, x_2, \cdots, x_n)$ can be factored in the form $f(x_1, x_2, \cdots, x_n) = f_1(x_1)f_2(x_2)\cdots f_n(x_n)$, where $f_i(x_i)$ is the frequency function of x_i, then the random variables $x_1, x_2, \cdots, x_n$ are said to be independently distributed.*

As an illustration, suppose that the number of automobile accidents, x, occurring in a given city in a given month possesses the frequency function $f(x) = e^{-m}m^x/x!$, where m is a positive constant. If y denotes the number of accidents in the following month, and if it possesses the same frequency function as x, and if x and y are independently distributed, then

$$f(x, y) = \frac{e^{-m}m^x}{x!}\frac{e^{-m}m^y}{y!} = \frac{e^{-2m}m^{x+y}}{x!y!}$$

However, if the function $f(x, y) = cm^{xy}/x!y!$ had been selected, then x and y would not be independent variables, because it is not possible to write this function as the product of a function of x alone and a function of y alone.

2.12 Continuous Frequency Functions

Thus far the discussion of probability has been confined to finite sample spaces in which all sample points are assigned the same probability. This simplification made it possible to derive the fundamental rules of probability by merely counting sample points. It will be assumed hereafter that these rules may also be applied to sample spaces for which the probabilities assigned to the sample points are not equal and in which there may be an infinite number of discrete sample points. As an illustration of a problem for which this extension of the

applicability of the rules of probability is needed, consider the problem of calculating the probability that the first head obtained in tossing a coin repeatedly will occur on or before the fourth toss. Here the sample space might conveniently consist of the infinite number of sample points represented by the infinite sequence of outcomes H, TH, TTH, $TTTH$, $\cdots$. If it is assumed that the coin is not biased, the probabilities that would be assigned to these sample points are $\frac{1}{2}$, $(\frac{1}{2})^2$, $(\frac{1}{2})^3$, $(\frac{1}{2})^4$, $\cdots$. It will be observed that the sum of these sample point probabilities is 1, as it should be. The random variable x here is a variable that can assume any one of the values 1, 2, 3, $\cdots$, and the problem is to calculate the value of $F(4)$.

The random variable of the preceding illustration is an example of what is called a discrete variable. A *discrete random variable* is a random variable that can assume only a finite number, or an infinite sequence, of distinct values. This means that the values can be arranged in a definite order.

Although the extension of the applicability of the rules of probability as indicated above enables one to consider a much larger class of problems than before, there are many important classes of problems that are still not covered. These problems involve sample spaces that contain all the points in an interval or intervals, rather than just a discrete set of points. For example, suppose an experiment consists in the weighing of an adult male from the population of a given city. Although there are only a finite number of individuals in the city and hence only a finite number of possible outcomes of the experiment, the mathematical model for such an experiment is much simpler if one conceives of an infinite number of individuals and conceives of all possible weights in some interval as being possible outcomes of the experiment. If the random variable x denoting the weight of an individual is introduced, then this assumes that x can take on any value in the interval, say, from 150 to 160 pounds. A random variable that can assume any value in some interval or intervals is called a *continuous random variable*. Such variables as weights, lengths, temperatures, and velocities, which are essentially variables involving measurement, are considered to be continuous variables. Although there are variables that are a mixture of the discrete and continuous types, the important problems in statistics usually involve either one or the other type of variable; hence, only these 2 distinct types will be considered.

For the purpose of discussing properties of continuous variables, consider a particular continuous random variable x that represents the thickness of a metal washer obtained from a certain machine turning

out washers. If the machine were permitted to turn out, say, 100 washers, and if the thicknesses of these 100 washers were measured to the nearest .001 inch, there would be available 100 values of x with which to study the behavior of the machine. If these 100 values were collected and represented in table form, one might find a table of values such as that displayed in Table 1, giving the absolute frequency f with which various values of x occurred. The word "frequency" usually implies the ratio of the observed number of values of x to the total number of observational values; however, it is also used to denote the numerator of this ratio. Throughout the subsequent chapters, if there is any question as to which meaning is being used, the words "relative frequency" and "absolute frequency" will be employed. In Table 1, absolute frequenices are recorded.

TABLE 1

x	.231	.232	.233	.234	.235	.236	.237	.238	.239
f	1	2	8	18	28	24	13	4	2

For the purpose of displaying these results graphically, a type of graph called a *histogram* will be used. A histogram is a graph of the type shown in Fig. 6, in which areas are used to represent observed

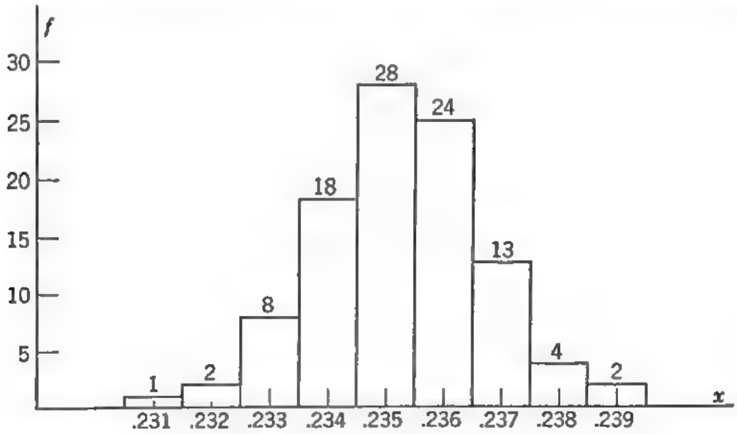

FIG. 6. Histogram for Table 1.

frequencies, particularly relative frequencies. Thus, the area of the rectangle that is centered at $x = .234$ should equal the relative frequency .18; however, in practice it is customary to choose any convenient unit on the y axis, with the result that the areas of the

rectangles may be only proportional to the corresponding frequencies rather than equal to them. The histogram shown in Fig. 6 for the data of Table 1 has been constructed with such a convenient choice of units; hence, areas there are only proportional to frequencies.

If the histogram is to be constructed so that areas will be equal to relative frequencies, then the total area of the histogram must equal 1 because the sum of the relative frequencies must equal 1. If h denotes the distance between consecutive x values, the height of the rectangle centered at, say, x_i will be f_i/Nh, where f_i denotes the absolute frequency of x_i. This result is obvious when it is realized that this ordinate when multiplied by the base h must equal the relative frequency f_i/N.

The histogram of Fig. 6 indicates the frequency with which various values of x were obtained for 100 runs of the experiment. If 200 runs had been made, the resulting histogram would have been twice as large as that based on 100 runs. In order to compare histograms based on different numbers of experiments, it is necessary to choose units on the y axis, as discussed in the preceding paragraph, in such a manner that the area of the histogram will always be equal to one. With this choice of units, the histogram would be expected to approach a fixed histogram as the number of runs of the experiment is increased indefinitely. Furthermore, if it is assumed that x can be measured as accurately as desired so that the unit on the x axis, h, can be made as small as desired, then the histogram would be expected to smooth out and approximate a continuous curve as the number of runs of the experiment is increased indefinitely and h is chosen very small. Such a curve is thought of as an idealization for the relative frequency with which different values of x would be expected to be obtained for runs of the actual experiment.

When the area of the histogram is made equal to 1, it follows from the preceding discussions that the sum of the areas of several neighboring rectangles is equal to the relative frequency with which the value of x was observed to lie in the interval that forms the base of those rectangles. Since this property will continue to hold as the number of runs of the experiment increases indefinitely, the area under the expected limiting, or idealized, curve between any 2 given values of x should be equal to the relative frequency with which x would be expected to lie in the interval determined by those values of x. The function $f(x)$ whose graph is conceived as being the limiting form of the histogram is treated as the mathematical model for the continuous random variable x and is called the frequency function of the variable. Since frequency in the case of a histogram is replaced by probability

in the case of a mathematical model, the definition of a frequency function for a continuous variable may be stated in the following form.

(21) DEFINITION: *A frequency function for a continuous random variable x is a function $f(x)$ that possesses the following properties:*

(*i*) $$f(x) \geq 0$$

(*ii*) $$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

(*iii*) $$\int_{a}^{b} f(x)\, dx = P\{a < x < b\}$$

where a and b are any two values of x, with $a < b$.

Property (i) is obviously necessary since negative probability has no meaning. Property (ii) corresponds to the requirement that the probability of an event that is certain to occur should be equal to one. Here x is certain to assume some real value when an observation of it is made. Although x is certain to assume some value, the probability that it will assume a stated value is 0 for a continuous random variable. If the range of x is not the entire real line, it is assumed that $f(x)$ is defined to be equal to 0 for those values outside the range of the variable.

As an illustration, consider the possibility of using $f(x) = ke^{-x}$ as a frequency function for x where k is some constant. From (i) it is clear that k must be positive. Since the integral of e^{-x} from $-\infty$ to $+\infty$ is infinite, it follows that the range of x must be restricted; hence assume, for example, that x can take on only non-negative values. Then $f(x)$ will be defined to be 0 for negative values and to be given by the formula for non-negative values. From (ii) it then follows that k must be equal to 1. The calculation of, say, $P\{1 < x < 2\}$ would then become

$$\int_{1}^{2} e^{-x} dx = e^{-1} - e^{-2} = .23$$

The graph of this frequency function and the representation of $P\{1 < x < 2\}$ as an area is given in Fig. 7.

Although $f(x)$ may be chosen at will in any given problem, a choice for which the resulting probabilities are not approximated well by observed relative frequencies is not likely to be a useful choice. As in the case of discrete variables, there are a number of particular frequency functions that have proved very useful in statistical work and whose explicit formulas will be considered later.

The frequency function for a continuous variable is often called the probability density function, or density function, of the variable; however, it is very convenient, and is becoming increasingly common, to use only the single name "frequency function" for both discrete and continuous variables.

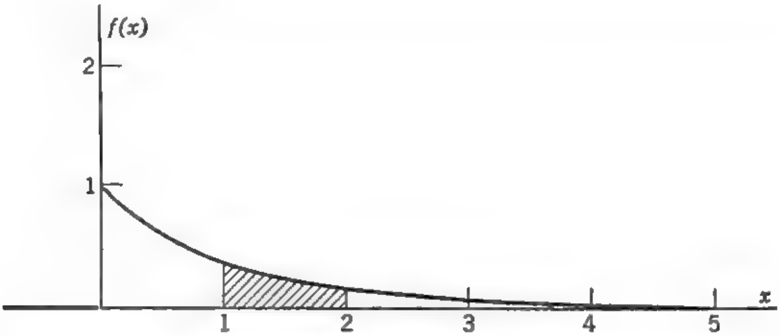

Fig. 7. Graph of a frequency function for a continuous variable.

The *distribution function*, $F(x)$, for the continuous variable x is defined by

$$F(x) = \int_{-\infty}^{x} f(t)\, dt \tag{22}$$

The graph of $F(x)$ for the preceding illustration is given in Fig. 8. It should be noted that $P\{1 < x < 2\}$ is now given by $F(2) - F(1)$,

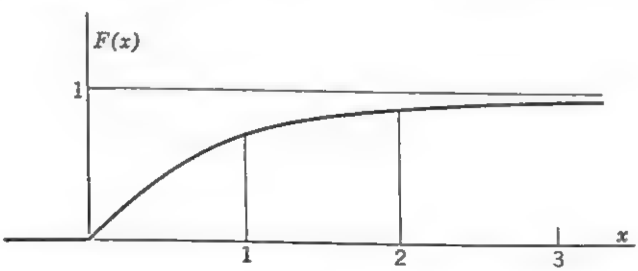

Fig. 8. Graph of the distribution function for a continuous variable.

that is, by the difference of the ordinates on the graph of $F(x)$. Here the graph was constructed by first determining $F(x)$ from definition (22). Thus

$$\begin{aligned} F(x) &= \int_0^x e^{-t}\, dt = 1 - e^{-x}, \quad x \geq 0 \\ &= 0, \quad x < 0 \end{aligned}$$

2.13 Joint Continuous Frequency Functions

A frequency function for several variables is a straightforward generalization of a frequency function for 1 variable. Thus, a frequency function for 2 variables x and y would be denoted by $f(x, y)$ and would be represented geometrically by a surface in 3 dimensions, just as a frequency function of 1 variable $f(x)$ was represented by a curve in 2 dimensions. The volume under the surface lying above the rectangle determined by $a < x < b$ and $c < y < d$ would give the probability that the random variables x and y will assume values corresponding to points lying inside this rectangle. The essential properties for a frequency function of several variables may be formalized as follows.

(23) DEFINITION: *A frequency function for n continuous random variables $x_1, x_2, \cdots, x_n$ is a function $f(x_1, x_2, \cdots, x_n)$ that possesses the following properties:*

(*i*) $f(x_1, x_2, \cdots, x_n) \geq 0$

(*ii*) $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \cdots, x_n)\, dx_1\, dx_2 \cdots dx_n = 1$

(*iii*) $\int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f(x_1, x_2, \cdots, x_n)\, dx_1\, dx_2 \cdots dx_n$

$$= P\{a_1 < x_1 < b_n, \cdots, a_n < x_n < b_n\}$$

As an illustration, consider the function $f(x, y) = e^{-(x+y)}$, which is a 2-dimensional generalization of the example used in the preceding section. If $f(x, y)$ is defined to be zero for negative values of x and y, it will be observed that (i) and (ii) will be satisfied. From (iii), the calculation of, say, $P\{1 < x < 2, 0 < y < 2\}$ will then be given by

$$\int_0^2 \int_1^2 e^{-x}e^{-y}\, dx\, dy = (e^{-1} - e^{-2})(e^0 - e^{-2}) = .20$$

The graph of $f(x, y)$ and the representation of $P\{1 < x < 2, 0 < y < 2\}$ as a volume is given in Fig. 9.

Continuous random variables that are unrelated in a probability sense are said to be independently distributed, just as in the case of discrete random variables. To say that continuous random variables are unrelated in a probability sense means that the probability that one of the variables will assume a value in a given interval is independent of what values the other variables assume. In order that this property shall hold, it suffices to define independence here exactly

as it was done for discrete variables; hence, definition (20) applies to continuous variables also. For the purpose of showing that the desired property holds, let $f(x_1, x_2, \cdots, x_n)$ be a frequency function satisfying (20). Then property (iii) of (23) implies that

$$\begin{aligned} P\{a_1 < x_1 < b_1, \cdots, a_n < x_n < b_n\} \\ &= \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f_1(x_1) f_2(x_2) \cdots f_n(x_n)\, dx_1\, dx_2 \cdots dx_n \\ &= \int_{a_1}^{b_1} f_1(x_1)\, dx_1 \int_{a_2}^{b_2} f_2(x_2)\, dx_2 \cdots \int_{a_n}^{b_n} f_n(x_n)\, dx_n \\ &= P\{a_1 < x_1 < b_1\} P\{a_2 < x_2 < b_2\} \cdots P\{a_n < x_n < b_n\} \end{aligned}$$

This result states that the probability that the variables $x_1, \cdots, x_n$ will simultaneously satisfy the indicated inequalities is equal to the product of the probabilities of the individual variables satisfying these inequalities. This property is the analogue for continuous variables of property (10) for events.

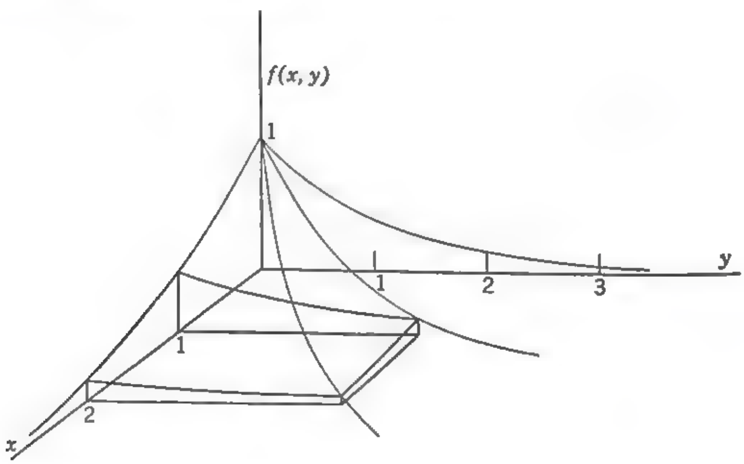

FIG. 9. Graph of a joint frequency function for two continuous variables.

The frequency function whose graph is given in Fig. 9 is an illustration of a joint frequency function of 2 independent random variables. In the present notation, $f_1(x_1) = e^{-x_1}$ and $f_2(x_2) = e^{-x_2}$.

It should be noted that in writing probability statements for continuous variables, such as in (23) (iii), it is irrelevant whether one uses $a_i < x_i < b_i$ or $a_i \leq x_i \leq b_i$ to determine the desired region. This would not be true, however, for discrete variables.

REFERENCES

A more extensive treatment of the various ideas and definitions of this chapter may be found in the following two books. Feller, W., *An Introduction to Probability Theory and its Applications*, John Wiley & Sons, New York. Neyman, J., *First Course in Probability and Statistics*, Henry Holt and Co., New York.

EXERCISES

1. A die has 2 of its sides painted red, 2 black, and 2 yellow. If the die is rolled twice, describe a 2-dimensional sample space for the experiment. What probabilities would you assign to the various sample points?

2. A coin is tossed 3 times. Describe a 1-dimensional sample space for the experiment. What probabilities would you assign to the various sample points?

3. If the die in Prob. 1 is rolled until a red side comes up, describe a sample space for the experiment. What probabilities would you assign to the various sample points?

4. Two balls are drawn from an urn containing 2 white, 3 black, and 4 green balls. (*a*) What is the probability that the first is white and the second is black? (*b*) What is this probability if the first ball is replaced before the second drawing?

5. One urn contains 2 white and 2 black balls; a second urn contains 2 white and 4 black balls. (*a*) If 1 ball is chosen from each urn, what is the probability that they will be the same color? (*b*) If an urn is selected at random and one ball is drawn from it, what is the probability that it will be a white ball? (*c*) If an urn is selected at random and 2 balls are drawn from it, what is the probability that they will be the same color?

6. Compare the chances of rolling a 4 with 1 die and rolling a total of 8 with 2 dice.

7. If 6 dice are rolled, what is the probability that each of the numbers from 1 through 6 will occur?

8. Assuming that the ratio of male children is $\frac{1}{2}$, find the probability that in a family of 6 children (*a*) all children will be of the same sex; (*b*) the 4 oldest children will be boys and the 2 youngest will be girls; (*c*) exactly half the children will be boys.

9. A, B, and C in order toss a coin. The first one to throw a head wins. What are their respective chances of winning? Note that the game may continue indefinitely.

10. Fourteen quarters and 1 five-dollar gold piece are in 1 purse, and 15 quarters are in another purse. Ten coins are taken from the first purse and placed in the second, and then 10 coins are taken from the second and placed in the first. How much money could you expect to get if you chose the first purse? How much if you chose the second purse?

11. If a poker hand of 5 cards is drawn from a deck, what is the probability that it will contain 2 aces?

12. What is the probability that a bridge hand will contain 13 cards of the same suit?

13. If a box contains 40 good and 10 defective fuses, and if 10 fuses are selected, what is the probability that they will all be good?

14. From a group of 50 people, 3 are to be chosen. Find the probability that none of 10 certain people in the group will be chosen.

15. If the numbers 1, 2, $\cdots$, n are arranged in a random order, what is the probability that the digits 1 and 2 appear next to each other?

16. What is the probability that the bridge hands of north and south together contain exactly 3 aces?

17. If a bridge player and his partner have 9 spades between them, what is the probability that the 4 spades held by the opponents will be split two and two?

18. Show that $\binom{n}{r-1} + \binom{n}{r} = \binom{n+1}{r}$.

19. Given the discrete frequency function $f(x) = e^{-1}/x!$, $x = 0, 1, 2, \cdots$, (*a*) calculate $P\{x = 2\}$; (*b*) calculate $P\{x < 2\}$; and (*c*) show that e^{-1} is the proper constant for this frequency function.

20. A coin is tossed until a head appears. (*a*) What is the probability that a head will first appear on the third toss? (*b*) What is the probability $f(x)$ that x tosses will be required to produce a head? (*c*) Graph the frequency function $f(x)$.

21. If the probability is $\frac{1}{2}$ that a finesse in bridge will be successful, (*a*) what is the probability that 3 out of 5 such finesses will be successful? (*b*) what is the probability, $f(x)$, that x out of 5 such finesses will be successful? (*c*) Graph the frequency function $f(x)$.

22. Graph the distribution function $F(x)$ for the frequency function obtained in Prob. 20.

23. Graph the distribution function $F(x)$ for the frequency function obtained in Prob. 21.

24. Six dice are rolled. Let x denote the number of ones and y the number of twos that show. Find an expression for $f(x, y)$, the probability of obtaining x ones and y twos.

25. Five cards are drawn from a deck. Let x denote the number of aces and y denote the number of kings that show. Find an expression for $f(x, y)$, the probability of obtaining x aces and y kings.

26. Given the continuous frequency function $f(x) = cxe^{-x}$, $x \geq 0$, (*a*) determine the proper value for c; (*b*) calculate $P\{x < 1\}$; and (*c*) calculate $P\{1 < x < 3\}$.

27. Given the continuous frequency function $f(x) = c$, $0 \leq x \leq 2$, (*a*) determine the proper value for c; (*b*) calculate $P\{x < 1\}$; and (*c*) calculate $P\{x > 1.5\}$.

28. Find the distribution function $F(x)$, and graph it if the frequency function of x is (*a*) $f(x) = 1$, $0 \leq x \leq 1$; (*b*) $f(x) = x$ for $0 \leq x \leq 1$ and $f(x) = -x + 2$ for $1 < x \leq 2$; and (*c*) $f(x) = [\pi(1 + x^2)]^{-1}$.

29. Given the joint frequency function $f(x, y) = xye^{-(x+y)}$, $x \geq 0$, $y \geq 0$, calculate $P\{x < 1, y < 1\}$.

30. Given the joint frequency function $f(x, y) = 8xy$, $0 \leq x \leq 1$, $0 \leq y \leq x$, calculate (*a*) $P\{x < .5, y < .25\}$; (*b*) $P\{x < .5\}$; and (*c*) $P\{y < .25\}$. (*d*) From the preceding calculations, what conclusions can be made concerning the independence of the variables x and y?

CHAPTER 3

Nature of Statistical Methods

3.1 Mathematical Models

The preceding two chapters have indicated to some extent the nature of statistical methods. The emphasis there was on experiments of the repetitive type, whether real or conceptual. Statisticians are mainly interested in constructing and applying mathematical models for experiments of this type. The advantage of such a model is that it enables the statistician to study properties of the experiment and to make predictions about the outcomes of future trials of the experiment, both of which would be difficult or impossible to do without such a model.

The process of constructing a model on the basis of experimental data and drawing conclusions from it is an example of *inductive inference.* When it is applied to statistical problems, it is usually called *statistical inference.* Thus, statisticians are principally engaged in making statistical inferences.

Most often the statistician is interested in constructing a mathematical model for a random variable associated with an experiment rather than for the experiment itself. For example, if x represents the number of defective parts that will be found in a lot of 100 parts submitted for inspection, he would prefer to have a model that predicts the frequency with which the various values of x will be obtained, rather than a model that predicts the frequency with which the various possible experimental outcomes will occur when 100 parts are selected from the production process. As a consequence, most of the models chosen by statisticians are frequency functions of random variables. Statistical inferences are therefore usually inferences about frequency functions of random variables.

As an illustration of the preceding ideas, suppose a biologist has observed that 44 out of 200 insects of a given type possess markings that are different from those of the rest. Suppose, further, that the biologist suspects that the markings are inherited according to a law

which implies that 25 per cent of such insects would be expected to possess the less common markings. If he assumes that the inheritance law is operating here and lets x represent the number of insects out of 200 that will possess the less common markings, then the model that he would naturally select is the frequency function

$$f(x) = \frac{200!}{x!(200-x)!}\left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{200-x} \tag{1}$$

This particular frequency function is the same as the frequency function given by (16), Chapter 2, because the two problems are equivalent from a probability point of view if the observations made on insects are considered as independent trials of an experiment.

If there had been no theory to suggest that ¼ of such insects should possess the unusual markings, the biologist might have chosen this same frequency function with the probability ¼ replaced by the observed relative frequency .22.

By means of (1) it would be possible for the biologist to make predictions about future sets of 200 observations and thus detect disagreements with his theory.

In its most general formulation, statistical inference is a type of decision making based upon probability. The statistician is largely engaged in constructing methods for making decisions. In a more limited sense, however, a large share of the inferences, or decisions, made by statisticians fall into 1 of 2 categories. Either they involve the testing of some hypothesis about the frequency function selected as the model, or they involve the estimation of parameters, or other characteristics, of this frequency function. These 2 types of statistical inference will be studied briefly in the next 2 sections from a general point of view, but will be applied throughout the book and studied further in Chapter 10.

3.2 Testing Hypotheses

Since the variety of statistical hypotheses that occur in applications is very large, a fairly general definition of what constitutes a statistical hypothesis is needed. Such a definition is the following.

(2) DEFINITION: *A statistical hypothesis is an assumption about the frequency function of a random variable.*

As an illustration for a discrete variable, consider the problem of the preceding section. If p denotes the proportion of all insects possessing the less common markings, then the assumption that $p = ¼$ is a statistical hypothesis. As an illustration for a continuous variable,

suppose the random variable x represents the time that elapses between two successive trippings of a Geiger counter in studying cosmic radiation and suppose it is assumed that the frequency function for x is a function of the form

$$f(x;\theta) = \theta e^{-\theta x} \tag{3}$$

where θ is a parameter whose value depends upon the experimental conditions. The assumption that the frequency function is a function of this particular form is obviously a statistical hypothesis. If it is assumed that the parameter θ is equal to 2, then this assumption is also a statistical hypothesis.

Now consider what is meant by testing a statistical hypothesis. A general definition can be expressed in the following form.

(4) DEFINITION: *A test of a statistical hypothesis is a procedure for deciding whether to accept or reject the hypothesis.*

This definition permits the statistician unlimited freedom in designing a test; however, he will obviously be guided by desirable properties of tests in designing them. Thus, a simple but ordinarily useless test is one in which a coin is tossed, and it is agreed to accept the hypothesis in question if, and only if, the coin turns up a head.

In order to illustrate how the statistician proceeds in attempting to design a test that possesses desirable properties, consider a problem related to the frequency function (3). Suppose a physicist is certain, from theoretical or experimental considerations, that the time that elapses between two successive trippings on a counter possesses the frequency function (3). Suppose further that he is quite certain that for the material with which he is working the value of the parameter is either 2 or 1, with his intuition favoring the value 2. In order to assist him in making a choice, the statistician might proceed in the following manner.

Assume that the frequency function (3) applies. Although this assumption constitutes a statistical hypothesis, it will not be tested here because the physicist is quite certain of the validity of this assumption. Assume that the parameter θ has the value 2. This assumption is the statistical hypothesis that will be tested. Denote this hypothesis by H_0. Let H_1 denote the alternative hypothesis that $\theta = 1$. Thus, the problem is one of testing the hypothesis H_0 against the single alternative H_1.

In order to test H_0, a single observation will be made on the random variable x; that is, a single time interval between 2 successive trippings of the counter will be measured. In real-life problems one usually